

Haoyang Wen, Ying Lin, Tuan Lai, Xiaoman Pan, Sha Li, Xudong Lin, Ben Zhou, Manling Li, Haoyu Wang, Hongming Zhang, Xiaodong Yu, Alexander Dong, Zhenhailong Wang, Yi Fung, Piyush Mishra, Qing Lyu, Dídac Surís, Brian Chen, Susan Windisch Brown, Martha Palmer, Chris Callison-Burch, Carl Vondrick, Jiawei Han, Dan Roth, Shih-Fu Chang, and Heng Ji. 2021. **RESIN: A dockerized schema-guided cross-document cross-lingual cross-media information extraction and event tracking system**. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 133–143, Online. Association for Computational Linguistics.

Chengbiao Yang, Weizhuo Li, Xiaoping Zhang, Runshun Zhang, and Guilin Qi. 2020. A temporal semantic search system for traditional Chinese medicine based on temporal knowledge graphs. In *Semantic Technology: 9th Joint International Conference, JIST 2019, Hangzhou, China, November 25–27, 2019, Revised Selected Papers 9*, pages 13–20. Springer.

Seid Muhib Yimam, Iryna Gurevych, Richard Eckart de Castilho, and Chris Biemann. 2013. **WebAnno: A flexible, web-based and visually supported system for distributed annotations**. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–6, Sofia, Bulgaria. Association for Computational Linguistics.

Ningyu Zhang, Xin Xu, Liankuan Tao, Haiyang Yu, Hongbin Ye, Xin Xie, Xiang Chen, Zhoubo Li, Lei Li, Xiaozhuan Liang, et al. 2022. DeepKE: A deep learning based knowledge extraction toolkit for knowledge base population. *arXiv preprint arXiv:2201.03335*.

Xuejiao Zhao, Huanhuan Chen, Zhenchang Xing, and Chunyan Miao. 2021. Brain-inspired search engine assistant based on knowledge graph. *IEEE Transactions on Neural Networks and Learning Systems*.

Xiaohan Zou. 2020. A survey on application of knowledge graph. In *Journal of Physics: Conference Series*, volume 1487, page 012016. IOP Publishing.

## A System Architecture

As shown in Fig. 3, we use modern full-stack framework MERN<sup>5</sup>, Docker and Python to build CollabKG. It consists of four components wrapped in a Docker container, namely web client, server, NLP server, and database. The NLP server deploys ChatIE to obtain automatic annotation results. In addition, the database is crucial, storing and managing information such as projects, texts, users, etc. This is achieved by maintaining three collections including Project, Text, and User. Project stores

<sup>5</sup>MongoDB-Express-React-Node <https://www.mongodb.com/mern-stack>

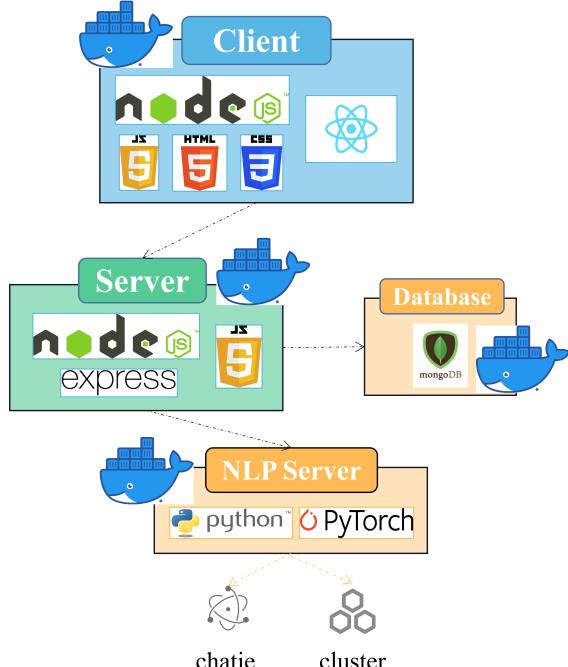


Figure 3: System Architecture.

the task details of the project, whether to perform the model update, semantic clustering, preprocessing, etc. Text manages samples, markups, etc. User stores the information of users like name and password.

## B Related Works

There are many existing open-source IE (*i.e.*, NER, RE, and EE) annotation tools for KG or EKG construction. We will describe them from several perspectives such as applicable tasks and annotation styles.

From the perspective of the three applicable tasks, FLAIR (Akbik et al., 2019), OpenNRE (Han et al., 2019), ODIN (Valenzuela-Escárcega et al., 2015), FitAnnotator (Li et al., 2021), WebAnno (Yimam et al., 2013), Redcoat (Stewart et al., 2019) and APLenty (Nghiem and Ananiadou, 2018) only focus on a single task. DeepKE (Zhang et al., 2022), Quickgraph (Bikaun et al., 2022), INCEPTION (Klie et al., 2018), TeamTat (Islamaj et al., 2020), TextAnnotator (Abrami et al., 2019), REES (Aone and Ramos-Santacruz, 2000) and SALKG (Tang et al., 2020) support two tasks. CoGIE (Jin et al., 2021), BRAT (Stenetorp et al., 2012), SLATE (Kummerfeld, 2019) and RESIN (Wen et al., 2021) support all three tasks. However, CoGIE and RESIN only support automatic labeling and require training, which is not suitable for low-resource scenarios where data is insufficient for

	P	R	F1	Time
<b>P1</b>	56.4	58.5	57.4	00:40:24
<b>P2</b>	70.3	60.4	65.0	00:42:21
<b>P3</b>	48.1	58.5	52.8	00:41:03
<b>P4</b>	71.4	66.0	68.6	00:42:07
<b>P5</b>	58.8	63.2	60.9	00:40:56
<b>P6</b>	79.4	80.2	79.8	00:40:31
<b>P7</b>	87.8	81.1	84.3	00:40:32
<b>P8</b>	73.3	72.6	73.0	00:41:17
<b>P9</b>	83.8	78.3	81.0	00:40:09
<b>P10</b>	75.9	77.4	76.6	00:41:40

Table 5: Human evaluation results on NER. **P** denotes participant.

training from scratch. BRAT and SLATE only support manual labeling. Moreover, BRAT sometimes is criticized for its difficulties in deployment (Neves and Ševa, 2019). SLATE is a command-line-based tool so it is not user-friendly.

From the perspective of the annotation styles (*i.e.*, automatic and manual labeling), INCEPTION, TextAnnotator, FitAnnotator, and APLenty support two annotation styles immediately. However, INCEPTION and TextAnnotator don’t support the learnability function (namely, self-renewal). Although FitAnnotator and APLenty utilize active learning to support two-way interaction, this is inflexible, not real-time, and requires a training process. In addition, as mentioned earlier, they only support a single task.

## C Detailed Results

To note, the standard labeling process includes multiple rounds of labeling, including validating and refining, iterating and improving. Validating and refining the annotations in the knowledge graph aims to ensure that the labels accurately represent the domain knowledge by double-checking. Iterating and improving denote continuously iterating and improving the knowledge graph based on feedback from multiple turns. In this work, all results are obtained in one round of annotation, so the numbers may seem to be low.

**NER** The results are presented in Tab. 5. The experimental group consisted of No. 6 to 10, while the control group consisted of No. 1 to 5. Before calculating the metrics we eliminated No.3 and No.8 because of their poor annotating quality.

**RE** The results are presented in Tab. 6. The experimental group consisted of No. 6 to 10. Before calculating the metrics we eliminated No.5 and No.6 because of their long annotation time or poor

	P	R	F1	Time
<b>P1</b>	57.0	47.6	51.9	01:29:42
<b>P2</b>	50.0	50.5	50.2	01:32:39
<b>P3</b>	35.3	39.8	37.4	01:23:44
<b>P4</b>	50.0	38.8	43.7	01:23:01
<b>P5</b>	47.9	33.0	39.1	02:15:39
<b>P6</b>	55.4	60.2	57.7	01:21:40
<b>P7</b>	68.1	62.1	65.0	01:14:04
<b>P8</b>	80.9	73.8	77.2	01:21:28
<b>P9</b>	75.3	73.8	74.5	01:23:56
<b>P10</b>	58.8	77.7	67.0	01:36:39

Table 6: Human evaluation results on RE. **P** denotes participant.

	P	R	F1	Time
<b>P1</b>	37.3/69.2	25.5/44.3	30.3/54.0	01:10:44
<b>P2</b>	43.4/75.0	30.0/49.2	35.5/59.4	01:27:25
<b>P3</b>	26.5/65.2	23.6/49.2	25.0/56.1	01:05:34
<b>P4</b>	36.6/72.5	30.9/47.6	33.5/57.4	01:07:34
<b>P5</b>	-	-	-	-
<b>P6</b>	37.7/84.6	41.8/72.1	39.7/77.9	01:04:51
<b>P7</b>	47.4/86.0	40.9/70.5	43.9/77.5	01:02:48
<b>P8</b>	44.9/82.7	48.2/70.5	46.5/76.1	01:33:12
<b>P9</b>	37.2/78.6	46.4/72.1	41.3/75.2	01:07:02
<b>P10</b>	44.0/86.3	46.4/72.1	45.1/78.6	01:02:21

Table 7: Human evaluation results on EE. The left and right numbers in Column F1 represent Arg-C and Trig-C, respectively. **P** denotes participant.

annotation quality. It is worth noting that since NYT11-HRL is obtained by remote supervision, the gold annotation does not cover all entities and relationships (Wei et al., 2019). Therefore, we re-examined and relabeled the 50 samples as the gold label.

**EE** The results are presented in Tab. 7. The experimental group consisted of No. 6 to 10. No. 5 participant was absent for personal reasons, so to align with the control group, we eliminated the results of No. 6 participant (poor annotation quality). It is worth noting that because there are so many tags in ACE05 (namely, 33), annotating them all would be too tricky. Therefore, we narrowed the tag range to 12 (within 50 samples).

## D Implementation of Unification

As shown in Fig. 4, to unify NER, IE, and EE tasks, we first observe their schemes and summarize the transformation rules among them. Then we uniformly model the schemes utilizing these rules and design an annotation format.

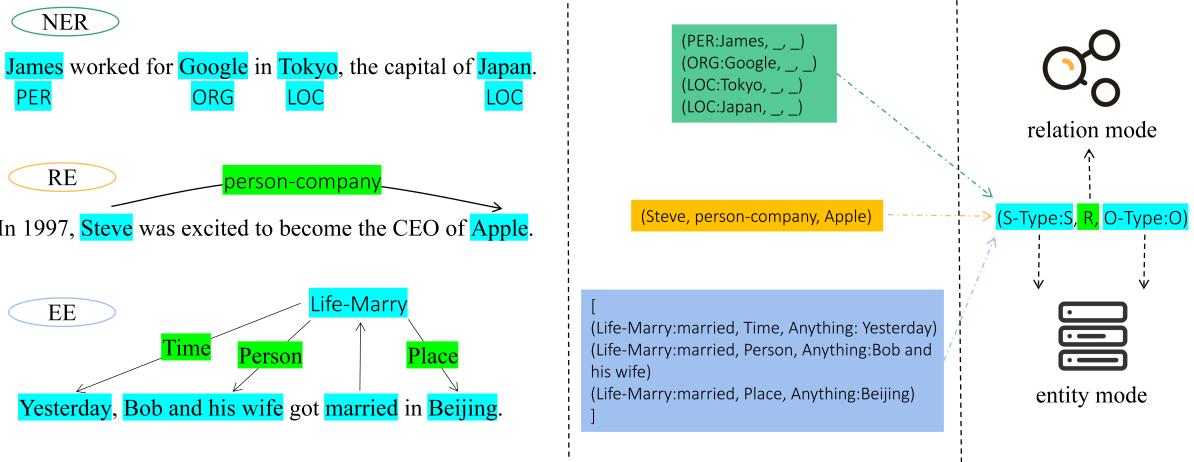


Figure 4: Illustration of unifying NER, IE, and EE tasks.

NER	[entity type 1, ..., entity type n]
RE	{relation type 1: [subject1, object1], ...}
EE	{event type 1: [argument role 1, ...], ...}

Table 8: Type list format for three IE tasks.

## E Automatic Labeling

We have adopted ChatIE<sup>6</sup> (Wei et al., 2023) as our approach for zero-shot information extraction, based on ChatGPT. ChatIE has shown impressive performance, even surpassing some full-shot models across various datasets. Its flexibility for customization is also a notable advantage because the type list allows customization (refer to Tab. 8).

Therefore, we have adopted ChatIE as the backbone of our automatic labeling module, with a slight modification that adds the trigger-related prompt template so that it can extract trigger words according to the event type. This allows for the extraction of trigger words based on the event type. For example, the following prompt template is used: When the event type of the given sentence above is “<event-type>”, please recognize the corresponding trigger word. The trigger word is the word or phrase that most clearly expresses event occurrences.\nOnly answer the trigger word, no extra word. The trigger word is:.

## F Human Evaluation Procedure

All participants were gathered in a conference room and asked to sign a consent form before being introduced to the task and annotation criteria. They

were then given model accounts to begin annotating without the use of external tools. Participants in the experimental group received ChatIE assistance. Once they finished annotating, they notified us using a communication tool, and we recorded the time spent on annotation. We evaluated metrics and conducted statistic analysis using a pre-written script.

<sup>6</sup><https://github.com/cocacola-lab/ChatIE>

## G Display

Our toolkit offers a range of features to enhance the user-friendly display. Firstly, it displays annotation progress and counts the results across multiple dimensions, such as entity, relation, and triples (see Fig. 5). Secondly, users can view a KG or EKG and filter the results (see Fig. 6 and 7). Thirdly, the tool provides a double-checking function for each text (see Fig. 8). Finally, our tool supports the filtering and exporting functions (see Fig. 9). Filter function includes saving, loading, quality filtering (accepted or suggested), and so on.

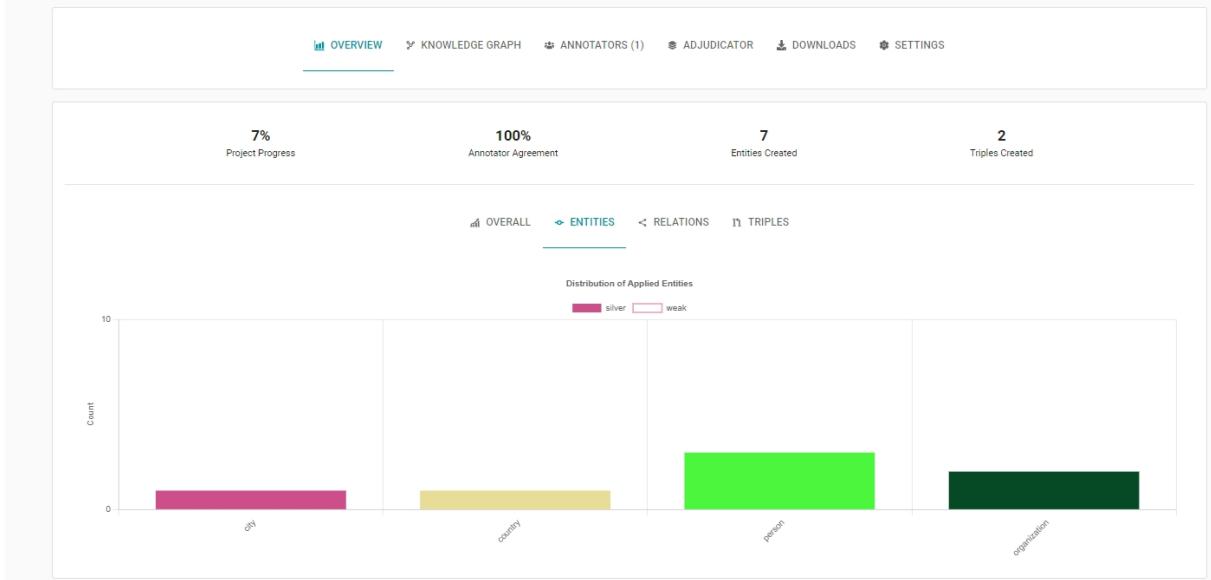


Figure 5: Overview of Dashboard.

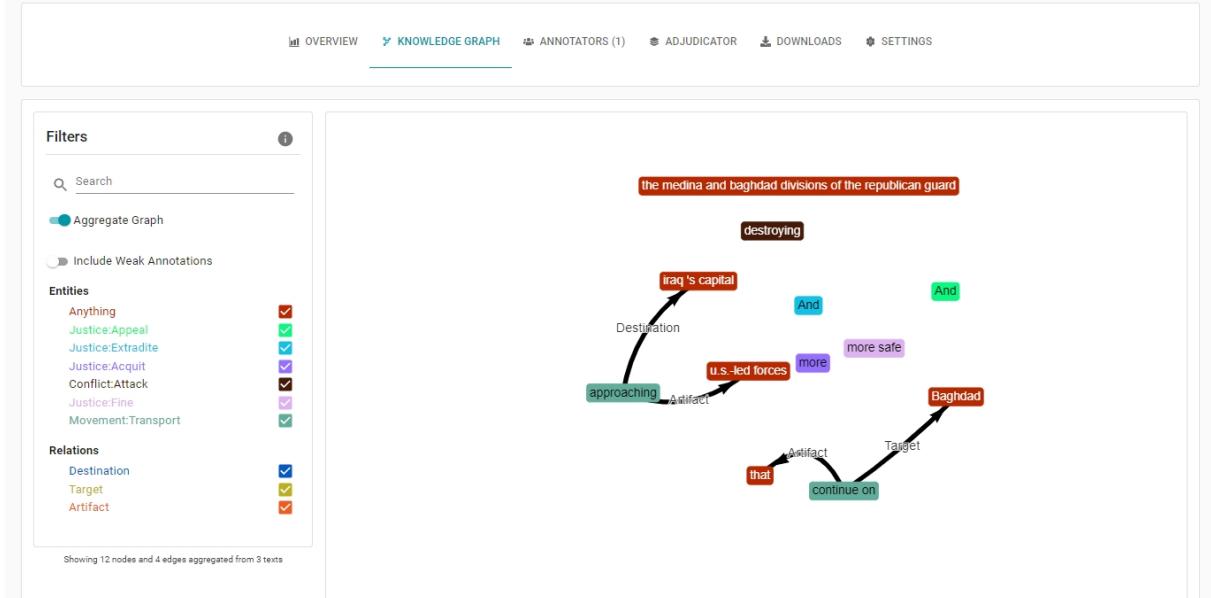


Figure 6: EKG display.

## H Project Creation Process

The project is created for each person on a dataset. The creation process can be divided into the following steps:

- *Configuration Setup:* Set up the details of the project: the name, the description, the configure of

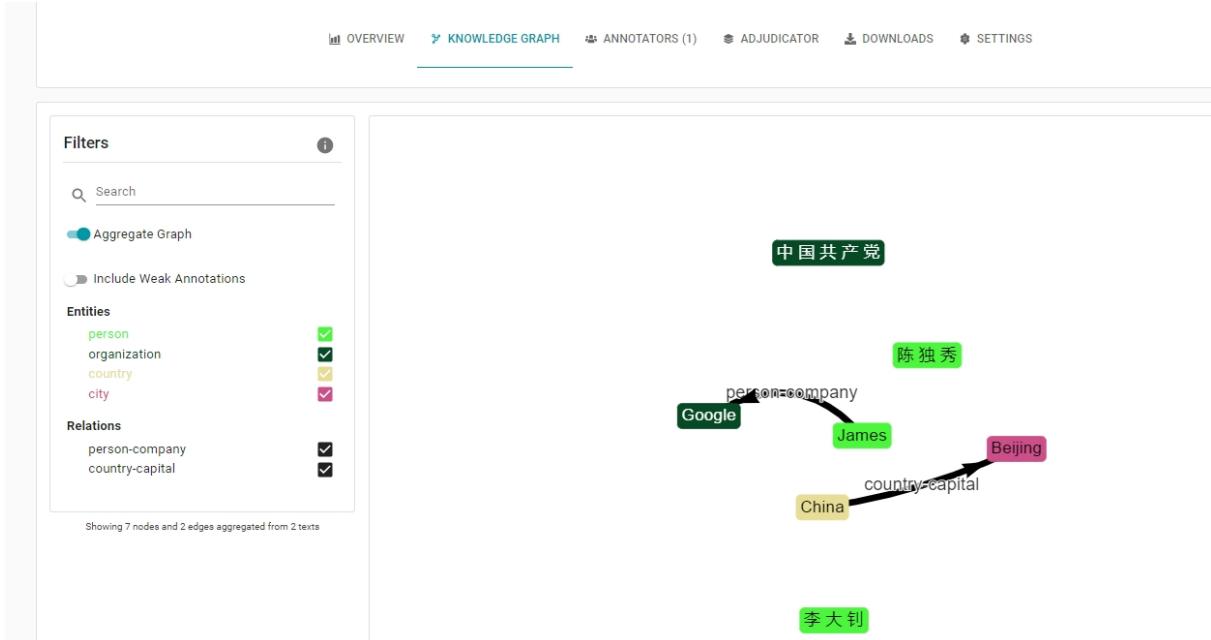


Figure 7: KG display.

The screenshot shows a double-checking display for the sentence "James worked for Google in Beijing, the capital of China.". It lists two annotations:

	Source	Relation	Target	Annotation
(P)	China Country	country-capital	Beijing City	capital
(D)	James Person	person-company	Google Organization	James worked for Google in Beijing ,

Annotations are highlighted in green. The overall statistics at the top right are 100% 100% 100% for overall entity relation.

Figure 8: Double-checking display.

The screenshot shows a download display with a sidebar for filters and a main table area. The filters include dropdowns for Type (Triples), Quality (Silver), Saved (Yes), Agreement (0), and Annotators To Include (Paper). A "FILTER" button is also present.

	Triples	Entities
P	2 2 Total Saved	7 4 7 0 Total Saved Silver Weak
G	2 2 Total Saved	7 4 7 0 Total Saved Gold Weak

A "DOWNLOAD TRIPLE ANNOTATIONS" button is located at the bottom right.

Figure 9: Download display.

multi-task (NER, RE, EE), whether to perform model update (see Sec.3.3) and text/document clustering features (Fig. 10). Clustering enables aggregation of texts with similar semantics (Fig. 2 A.IV) so that annotators can focus more on a certain class of concepts and thus increase productivity. Our toolkit implements cohesive clustering by encoding documents with SBERT (Reimers and Gurevych, 2019) sentence embeddings.

- *Uploading Data:* It supports keyboard input and uploading files (Fig. 11).
- *Pre-processing:* Pre-processing function includes character casing, specified-character removal, and text de-duplication (Fig. 12).
- *Scheme Setup:* Build an ontology/scheme for the current task. Users can choose from the preset ontology or **customize** their own scheme. For RE, the RELATION TYPES format is relation@[subject, object], where subject/object refers to head/tail entity type in triples (Fig. 13). For EE, the ENTITY TYPES is filled with a pseudo token (namely \_), and the RELATION TYPES format is event-type@[argument role 1, argument role 2, ...] (Fig. 14). It is worth noting that CollabKG will complete the processing to convert role to relation and event-type to entity-type on the back-end. Unlike other IE annotation tools, our tool supports **hierarchical labels** (Fig. 15) and **relation constraint** (Fig. 16). The hierarchical labels facilitate the management of complex schemes. The relation constraint is a predetermined that a relationship can only occur between certain entity types. Consequently, this feature can narrow down the annotator’s attention and improve the productivity and consistency of the annotators.
- *Preannotation:* Users can choose to upload pre-annotated entities and relations of the current corpus. This can reduce annotation effort by pre-applying tags based on external resources such as gazetteers (Fig. 17).
- *Review:* Summarize the current project. Hence users can check and make changes (Fig. 18).

Finally, when the user clicks the CREATE button, the project creation process is completed and will appear in the panel (Fig. 19).

The screenshot shows the 'New Project' interface with a dark header bar containing 'NEW PROJECT' and 'MENU'. Below the header, there's a 'Details' section with fields for 'Project Name' (set to 'RE') and 'Project Description' (set to 're present'). The 'Multi-task Configuration' section has a checked checkbox for 'Entity-Relation Triples Annotation'. The 'Model Update' section has an unchecked checkbox for 'Perform model update'. The 'Document Clustering' section has an unchecked checkbox for 'Perform document clustering'. On the right side, a vertical dashed line connects six numbered steps: 1. details, 2. upload, 3. preprocessing, 4. schema, 5. preannotation, and 6. review. At the bottom left is a 'BACK' button, and at the bottom right is a 'SAVE AND CONTINUE >' button.

Figure 10: Detail of project creation.

**New Project**

**Upload**  
Create Or Upload A Corpus, Note: Words (Including Punctuation) Should Be Separated By Spaces, Like "I Love Bob , Which Is My Father ."

**Project Corpus** 50 RE-ENG-SEED42.TXT

Project Corpus \*  
 Syria and Iran are with the opposition .  
 The obvious choice for that role is Senator Chuck Hagel of Nebraska , a decorated veteran who has been a consistent critic of the administration 's conduct of the war and who has not ruled out a run for the presidency .  
 In Nigeria , the golf resort in Uyo is temporarily closed .  
 Roberto González Echevarría , a professor of literature at Yale University and the author of " The Pride of Havana : A History of Cuban Baseball , " said he was informally advising Garagiola and had the impression that baseball officials wanted to work with Cuba .  
 A few years ago , after Mr. Yeltsin developed symptoms of further heart trouble , doctors at the Berlin Heart Center in Germany performed an angioplasty .  
 He turned for guidance to Daniel Pauly , director of the fisheries center at the University of British Columbia , which maintains an elaborate global database on fishing .

BACK SAVE AND CONTINUE >

Figure 11: Uploading of project creation.

**New Project**

**Preprocessing**  
Apply Text Preprocessing To Your Corpus

**Preprocessing Actions**  
 Lower Case  Remove Characters  Remove Duplicates  
 Characters To Remove  
 ~'?,!();[]\_-\$

50 → 50  
Corpus Size 897 → 863  
↓4%  
Vocab Size 1,791 → 1,594  
↓11%  
Token Size

**Corpus Preview**  
 syria and iran are with the opposition  
 The obvious choice for that role is senator chuck hagel of nebraska a decorated veteran who has been a consistent critic of the administration 's conduct of the war and who has not ruled out a run for the presidency  
 in nigeria the golf resort in uyo is temporarily closed  
 roberto gonzález echevarría a professor of literature at yale university and the author of " the pride of havana a history of cuban baseball " said he was informally advising garagiola and had the impression that baseball officials wanted to work with cuba  
 a few years ago after mr yeltsin developed symptoms of further heart trouble doctors at the berlin heart center in germany performed an angioplasty  
 he turned for guidance to daniel pauly director of the fisheries center at the university of british columbia which maintains an elaborate global database on fishing  
 john j mackowski insurance executive and proponent for corporate responsibility and ethical practice died of lung cancer february 24 at his home in ponte vedra beach florida

Figure 12: Preprocessing of project creation.

**New Project**

**Schema**  
Build An Ontology/Schema For Entity-Relation Triples Annotation. Ps: The Relation Types Format Is "Relation@[Subject, Object]", Where Subject/Object Refers To Head/Tail Entity Type In Triples."

ENTITY TYPES RELATION TYPES

Select Custom to build your own ontology or select a preset  
 Preset Ontologies  
 NYT11\_HRL

location-located\_in@{boca}  administrative\_division-cc  person-place\_lived@{pers}  person-company@{pers}  person-nationality@ {pers}

ADD BRANCH NODE CLEAR ALL

BACK SAVE AND CONTINUE >

Figure 13: Scheme setup for RE of project creation.

**New Project**

**Schema**  
Build An Ontology/Schema For Event Annotation. Ps: 1. The Event Types Only Choice Preset "EE", Where "Anything" Stands For Event Argument, Of Course You Can Change It For You Like. 2. The Relation Types Format Is "Event-Type@[Argument Role 1, Argument Role 2, ...]."

Entity Types      Relation Types

Select **Custom** to build your own ontology or select a preset  
Preset Ontologies  
ACE05\_EE

**SAVE AND CONTINUE >**

1. details  
2. upload  
3. preprocessing  
4. schema  
5. preannotation  
6. review

Figure 14: Scheme setup for EE of project creation.

**New Project**

**Schema**  
Build An Ontology/Schema For Entity Annotation

Entity Types

Select **Custom** to build your own ontology or select a preset  
Preset Ontologies  
Custom

**SAVE AND CONTINUE >**

1. details  
2. upload  
3. preprocessing  
4. schema  
5. preannotation  
6. review

Figure 15: Hierarchical labels of project creation.

**New Project**

Type In Triples.

Entity Types      Relation Types

Select **Custom** to build your own ontology or select a preset  
Preset Ontologies  
NYT11\_HRL

person-nationality@ [person, country]

Domain (2)	Range (2)
person	person
organization	organization
location	location
country	country
city	city

CLEAR    SELECT ALL    CLEAR    SELECT ALL

CANCEL    **SAVE**

**SAVE AND CONTINUE >**

2. upload  
3. preprocessing  
4. schema  
5. preannotation  
6. review

Figure 16: Relation constraint of project creation.

**New Project**

NEW PROJECT MENU ▾

**Preannotation**  
Upload Data For Pre-Annotation

Before commencing annotation with AutoKG, you can upload a dictionary for pre-annotation of entities and triples. The types specified in each dictionary must be consistent with the ontologies specified in this project. The dictionary items must be consistent with relation constraints (if applicable).

**Entity Preannotation**  
Upload a set of known entity annotations to pre-annotate your uploaded corpus.

Upload dictionary to pre-annotate entities **UPLOAD ENTITIES**

**Typed Triple Preannotation**  
Upload a set of known typed triple annotations to pre-annotate your uploaded corpus. Format: (source span, source type, relation type, target span, target type, offset).

Upload dictionary to pre-annotate entities with relations **UPLOAD TRIPLES**

**BACK** **SAVE AND CONTINUE >**

Figure 17: Preannotation of project creation.

**New Project**

NEW PROJECT MENU ▾

**Review**  
Review Project Before Creation

Step	Summary
Details	Name: RE   Description: re present   Task: Entity Typing and Closed Relation Extraction   No Model Update   No Clustering
Upload	50 Documents
Preprocessing	Lower Case   Removed Special Characters
Schema	person, organization, location, country, city   12 relation types created
Preannotation	No entity preannotation performed   No typed triple preannotation performed

Project creation may take a few minutes if your corpus is very large or semantic clustering is being performed

**BACK** **CREATE**

Figure 18: Review of project creation.

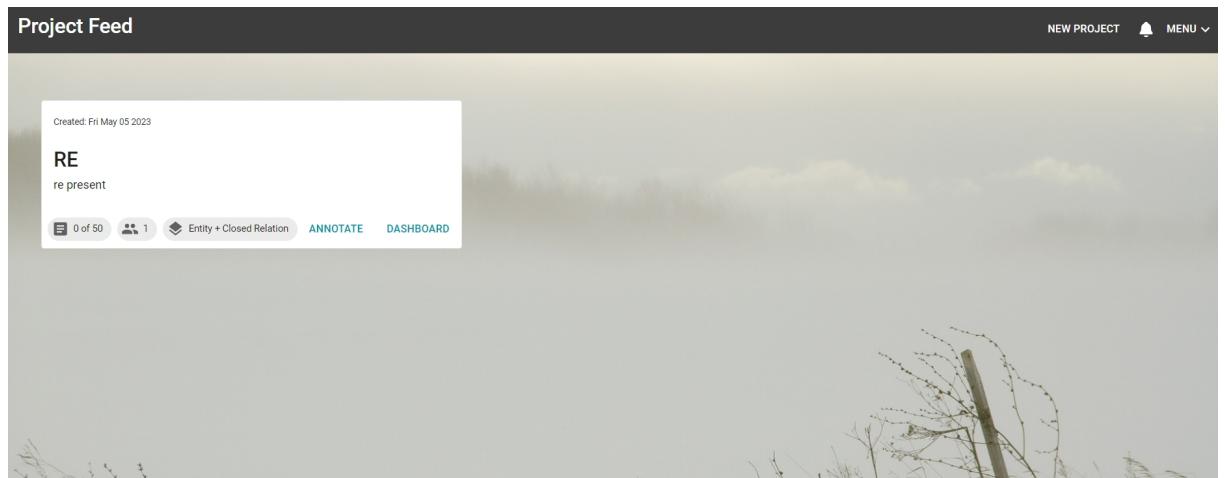


Figure 19: Feed of projects.